

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

**EP 0 875 567 A2**

(12)

**EUROPÄISCHE PATENTANMELDUNG**

(43) Veröffentlichungstag:

04.11.1998 Patentblatt 1998/45

(21) Anmeldenummer: 98106426.4

(22) Anmeldetag: 08.04.1998

(51) Int. Cl.<sup>6</sup>: **C12N 15/12**, C07K 14/47,

C12N 15/63, C12N 1/21,

G01N 33/68, C07K 16/18,

A61K 48/00

(84) Benannte Vertragsstaaten:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE

Benannte Erstreckungsstaaten:

AL LT LV MK RO SI

(30) Priorität: 30.04.1997 DE 19718249

(71) Anmelder:

**BASF AKTIENGESELLSCHAFT**  
67056 Ludwigshafen (DE)

(72) Erfinder:

• Peukert, Karen

35094 Lahntal-Sterzhausen (DE)

• Haenel, Frank, Dr.

07745 Jena (DE)

• Eilers, Martin, Prof. Dr.

35043 Marburg-Cappel (DE)

(54) **Myc-bindende Zinkfinger-Proteine, ihre Herstellung und ihre Verwendung**

(57) Neue Myc-bindende Zinkfingerproteine, ihre  
Herstellung und ihre Verwendung.

**EP 0 875 567 A2**

## Beschreibung

Die vorliegende Erfindung betrifft Myc-bindende Zinkfinger-Proteine, ihre Herstellung und ihre Verwendung.

Myc ist ein spezifisch an DNA bindendes Protein. Es wird zur Familie der Helix-Loop-Helix/Leucin-Zipper (HLH/LZ) Transkriptionsfaktoren gezählt (Landschulz et al, 1988, Murre et al., 1989). Myc ist ein zentraler Transkriptionsaktivator, der mit dem Protein Max (Amati et al., 1993) einen Komplex bildet und durch diesen molekularen Mechanismus andere Gene aktiviert, beispielsweise alpha-Prothymosin, Ornithindecaboxylase und odc25A.

Von Schulz et al, 1995, wurde ein 13 Zinkfinger enthaltendes Protein aus der Maus beschrieben, dessen zelluläre Funktion jedoch unklar ist.

Aufgrund seiner Schlüsselstellung in der Transkription bietet Myc einen Ansatzpunkt zum Verständnis von zellulären, insbesondere von pathophysiologischen Prozessen.

Es bestand daher die Aufgabe, weitere Informationen über die molekulare Wirkungsweise von Myc, insbesondere über die Myc vermittelte Genrepression bereitzustellen.

Gegenstand der Erfindung ist ein Protein mit der in SEQ ID NO:2 dargestellten Aminosäuresequenz. Dieses Protein besitzt dreizehn Zinkfingerdomänen.

Es weist folgende biologischen Eigenschaften auf:

- Spezifische Bindung an Myc,
- Transaktivierung des Adenovirus Major Late (AdML) Promotors,
- Transaktivierung des Cyclin D1 Promotors,
- durch Assoziation mit Myc wird die Transaktivierung gehemmt,

in Abwesenheit von Myc ist das Protein im wesentlichen im Cytosol assoziiert mit Mikrotubuli zu finden.

Ein weiterer Gegenstand der Erfindung sind Proteine, die sich aus der SEQ ID NO:2 dargestellten Struktur durch Substitution, Insertion oder Deletion von einem oder mehreren Aminosäuren ableiten lassen, wobei diese Proteine noch die wesentlichen biologischen Eigenschaften des durch SEQ ID NO:2 beschriebenen Proteins besitzen. Diese Proteine werden im folgenden Muteine genannt. Unter wesentlichen Eigenschaften wird die spezifische Bindung der Muteine an Myc verstanden.

Die oben aufgeführten Eigenschaften des durch SEQ ID NO:2 beschriebenen Proteins müssen nicht alle bei den Muteinen vorhanden sein, solange die spezifische Bindung an Myc gegeben ist. Bevorzugt sind jedoch diejenigen Muteine, die alle der oben aufgeführten Eigenschaften besitzen.

Die Anzahl der durch Insertion Substitution oder Deletion gegenüber dem durch SEQ ID NO:2 beschriebenen Protein veränderten Aminosäuren kann zwischen 1 und 100, bevorzugt zwischen 1 und 50 Aminosäuren variieren. Die Veränderungen können in einem kleineren Bereich des Moleküls konzentriert oder auch über das ganze Molekül verteilt sein.

Bevorzugte Veränderungen sind konservative Substitutionen, bei denen eine Aminosäure durch eine andere Aminosäure mit ähnlicher Raumerfüllung, Ladung oder Hydrophilie ersetzt wird.

Beispiele für solche konservativen Substitutionen sind

- Ersatz von Arg durch Lys oder umgekehrt,
- Ersatz von Arg durch His oder umgekehrt,
- Ersatz von Asp durch Glu oder umgekehrt,
- Ersatz von Asn durch Gln oder umgekehrt,
- Ersatz von Cys durch Met oder umgekehrt,
- Ersatz von Cys durch Ser oder umgekehrt,
- Ersatz von Gly durch Ala oder umgekehrt,
- Ersatz von Val durch Leu oder umgekehrt,
- Ersatz von Val durch Ile oder umgekehrt,
- Ersatz von Leu durch Ile oder umgekehrt,
- Ersatz von Phe durch Tyr oder umgekehrt,
- Ersatz von Phe durch Trp oder umgekehrt,
- Ersatz von Ser durch Thr oder umgekehrt.

Die Veränderungen können auch kombiniert werden, z.B. eine oder mehrere Substitutionen mit Deletionen und/oder Insertionen.

Ein weiterer Gegenstand der Erfindung sind Nukleinsäuresequenzen, die für die oben beschriebenen Proteine codieren. Solche Nukleinsäuresequenzen sind bevorzugt DNA, insbesondere cDNA Sequenzen, in einzelsträngiger oder doppelsträngiger Form.

Bevorzugte Nukleinsäuresequenzen sind solche mit der in SEQ ID NO:1 dargestellten Sequenz und solche, die mit dieser Sequenz einen hohen Verwandtschaftsgrad aufweisen, beispielsweise solche, die für das gleiche Protein codieren wie SEQ ID NO:1. Weitere bevorzugte Nukleinsäuresequenzen sind solche, die für ein Protein codieren, das 95% oder mehr Identität mit dem Protein der Sequenz SEQ ID NO:2 aufweist.

Ein weiterer Gegenstand der Erfindung sind Vektoren, die eine der oben beschriebenen Nukleinsäuresequenzen in funktioneller Verknüpfung mit einem oder mehreren Regulationselementen tragen. Unter Regulationselemente sind Nukleinsäurefragmente zu verstehen, die auf Transkription oder Translation einen regulierenden Einfluß haben, beispielsweise Promotoren, Enhancer, Polyadenylierungsstellen, ribosomale Bindungsstellen.

Die mit solchen Vektoren transformierten Wirtsorganismen sind ebenfalls Gegenstand der Erfindung. Als Wirtsorganismen geeignet sind Mikroorganismen, pflanzliche oder tierische Zellen oder Lebewesen. Bevorzugte Wirtsorganismen sind eukaryontische Zellen und Lebewesen. Der Begriff Wirtsorganismus umfaßt auch beispielsweise transgene Tiere und Pflanzen.

Die Herstellung der erfindungsgemäßen Proteine erfolgt bevorzugt mit Hilfe gentechnischer Verfahren. Ein Wirtsorganismus, der die Erbinformation für die erfindungsgemäßen Proteine trägt, wird unter Bedingungen kultiviert, die die Expression des Proteins erlauben. Diese Bedingungen - wie Temperatur, Nährmedium, Zelldichte - hängen weitgehend von der Wahl des Wirtsorganismus ab. Solche Bedingungen sind jedoch dem Fachmann für die einzelnen Wirtsorganismen geläufig.

Die exprimierten Proteine werden anschließend, ggf. nach Aufbrechen des Wirtsorganismus, vom Wirtsorganismus abgetrennt und in reiner Form durch bekannte Methoden der Proteinreinigung, wie Fällung, Chromatographie, Elektrophorese in reiner Form isoliert. Ein weiterer Gegenstand der Erfindung ist die Verwendung der Proteine als Antigen zur Herstellung von Antikörpern, sowie die so erhaltenen Antikörper. Es lassen sich durch dem Fachmann bekannte Verfahren polyklonale Antiseren oder auch monoklonale Antikörper herstellen.

Die erfindungsgemäßen Proteine eignen sich auch als Testsysteme zur Auffindung von potentiellen selektiven Transkriptionsmodulierenden Substanzen. Dies läßt sich besonders gut testen, indem man die Fähigkeit der Proteine, mit Myc einen Proteinkomplex zu bilden, ausnützt. Ein weiterer Gegenstand der Erfindung ist daher ein Verfahren zur Identifizierung von spezifischen transkriptionsmodulierenden Substanzen, das folgende Schritte umfaßt:

(a) Inkubation des Proteins gemäß Anspruch 1 mit dem Genprodukt von myc unter Bedingungen, unter denen sich ein Proteinkomplex zwischen diesen beiden Proteinen ausbildet,

(b) Inkubation der beiden Proteine unter ansonst gleichen Bedingungen wie (a) jedoch in Anwesenheit einer oder mehrerer Substanzen, die auf spezifische transkriptionsmodulierende Aktivitäten zu testen sind,

(c) Ermitteln des Unterschiedes in der Proteinkomplexbildung zwischen (b) und (a),

(d) Auswahl solcher Substanzen, bei denen gemäß Schritt (b) eine andere Proteinkomplexbildung erhalten wurde als bei Schritt (a).

Es lassen sich damit Substanzen auffinden, die die Proteinkomplexbildung zwischen den neuen Zinkfingerprotein und Myc fördern, aber auch solche, die sie unterbinden.

Die erfindungsgemäßen Nukleinsäuresequenzen eignen sich auch zur Gentherapie von Erkrankungen, bei denen die durch Myc vermittelte Transkription gestört ist.

Beispielsweise können zusätzliche Gensequenzen eingebracht werden um so die zelluläre Konzentration der Zinkfingerproteine zu erhöhen. Es kann aber auch gewünscht sein, daß die Konzentration der Zinkfingerproteine erniedrigt werden soll. In diesem Falle bietet sich eine Gentherapie auf antisense Basis an, wobei man eine zu dem Zinkfingerproteingen komplementäre Nukleinsäure oder Nukleinsäurederivat appliziert, und somit die Expression des Zinkfingerproteingens reduziert.

Die weitere Ausgestaltung der Erfindung ist in den folgenden Beispielen aufgeführt.

#### Beispiel 1

Isolierung der DNA mit der durch SEQ ID NO:1 beschriebenen Struktur

Vorausgegangene Arbeiten hatten gezeigt, daß die Integrität der Helix-Loop-Helix Domäne von Myc kritisch für die Genrepression durch Myc in stabilen Zelllinien war (Philipp et al., 1994). Um neue Proteine zu identifizieren, die mit dem C-Terminus von Myc interagieren, wurde ein DNA-Fragment, das für die basische Region und die HLH/LZ Domäne (Aminosäuren 355-439 des humanen Myc) codiert, im Leserahmen an die DNA bindende Domäne von GAL4 (Aminosäure 1-147) fusioniert und als Köder in einem "Two-Hybrid-Screen" (Fields and Song, 1989) benutzt.

$2 \times 10^5$  unabhängige Transformanten einer HeLa cDNA Bibliothek, markiert mit der GAL4 Aktivierungsdomäne, wurden gescreent. Ein Clon mit  $\beta$ -Galaktosidaseaktivität wurde weiter charakterisiert. Es wurde keine Interaktion zwischen dem von diesem Clon codierten Protein und der DNA Bindungsdomäne von GAL4 allein oder einer GAL4-BCY-1 Chimäre, die als Negativkontrolle benutzt wurde, festgestellt.

5 Die Interaktion mit Myc wurde aufgehoben durch Deletion der HLH-Domäne in Myc (370-412), nicht aber durch Insertion der vier Aminosäuren zwischen der HLH Domäne und dem Leucin-Zipper (In 412) oder durch Deletion des gesamten Leucin-Zippers (412-434). Eine spezifische Interaktion wurde auch nachgewiesen mit N-Myc aber keine mit MAX oder USF, zwei HLH-Proteinen, die mit Myc nahe verwandt sind.

10 cDNA-Moleküle mit voller Länge wurden durch ein 5'-RACE-Protokoll isoliert und sequenziert (SEQ ID NO:1). Sie codieren ein Protein mit 803 Aminosäuren (SEQ ID NO:2) mit einem theoretischen Molekulargewicht von 87,970 Dalton. Das Protein wurde Miz-1 für Myc-Interacting-Zincfinger-Protein-1 genannt.

Die Sequenzierung ergab, daß der isolierte Clon für ein Zinkfingerprotein mit 13 Zinkfingern codierte, 12 davon unmittelbar geclustert in der C-terminalen Hälfte des Proteins.

## 15 Beispiel 2

### Herstellung von Muteinen

20 Ausgehend von der in SEQ ID NO:1 dargestellten Nukleinsäuresequenz können mit dem Fachmann geläufigen Methoden der Gentechnik Nukleinsäuren hergestellt werden, die für veränderte Proteine (Muteine) codieren. Die Herstellung der Muteine selbst erfolgt zweckmäßigerweise durch Expression einer Nukleinsäure in einem geeigneten Wirtsorganismus.

## Beispiel 3

25

### Assoziation des Proteins SEQ ID NO:2 mit Myc

30 Der C-Terminus des Proteins SEQ ID NO:2 (Aminosäure 269-803) wurde mit der Glutathion-Transferase (GST) (Smith and Johnson, 1988) fusioniert, das GST-Miz-1 Fusionsprotein gereinigt und mit in vitro synthetisiertem, radioaktiv markiertem Myc Protein inkubiert. Myc assoziiert spezifisch mit GST-Miz-1, jedoch nicht mit GST. Eine Mutante von Myc, der die HLH Domäne fehlt, konnte nicht mit GST-Miz-1 assoziieren. Radioaktiv markiertes Max interagiert weder mit GST-Miz-1 noch mit GST. Jedoch kann mit Hilfe von Myc Max an GST-Miz-1-Kügelchen in vitro binden, was dafür spricht, daß Miz-1 und Max mit unterschiedlichen Flächen der HLH-Domäne von Myc interagieren.

## 35 Literaturverzeichnis

Amati, B., Brooks, M. W., Levy, N., Littlewood, T. D., Evan, G. I., and Land, H. (1993). Oncogenic activity of the c-Myc protein requires dimerization with Max. *Cell* 72, 233-245.

40 Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* 340, 245-246.

Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* 240, 1759-1764.

45 Murre, C., SchonleberMcCaw, P., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56, 777-783.

Philipp, A., Schneider, A., Väsrik, I., Finke, K., Xiong, Y., Beach, D., Alitalo, K., and Eilers, M. (1994). Repression of Cyclin D1: a Novel Function of MYC. *Mol. Cell. Biol.* 14, 4032-4043.

50

Schulz, T. C., Hopwood, B., Rathjen, P. D., and Wells, J. R. (1995). An unusual arrangement of 13 zinc fingers in the vertebrate gene Z13. *Biochem. J.* 311, 219-224.

55 Smith, D. B., and Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-S-transferase. *Gene* 67, 31-40.

## SEQUENZPROTOKOLL

## (1) ALGEMEINE INFORMATION:

5

## (i) ANMELDER:

- (A) NAME: BASF Aktiengesellschaft
- (B) STRASSE: Carl-Bosch-Strasse 38
- (C) ORT: Ludwigshafen
- (E) LAND: Bundesrepublik Deutschland
- (F) POSTLEITZAHL: D-67056
- (G) TELEPHON: 0621/6048526
- (H) TELEFAX: 0621/6043123
- (I) TELEX: 1762175170

10

15

- (ii) ANMELDETITEL: Myc-bindende Zinkfingerproteine
- (iii) ANZAHL DER SEQUENZEN: 2

## (iv) COMPUTER-LESBARE FORM:

- (A) DATENTRÄGER: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) BETRIEBSSYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPA)

20

## (2) INFORMATION ZU SEQ ID NO: 1:

25

## (i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 2680 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

30

## (ii) ART DES MOLEKÜLS: cDNS zu mRNS

## (iii) HYPOTHETISCH: NEIN

35

## (iii) ANTISENSE: NEIN

## (ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: 5'UTR
- (B) LAGE: 1..159

40

## (ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: CDS
- (B) LAGE: 160..2571

45

## (ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: 3'UTR
- (B) LAGE: 2572..2680

## (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 1:

50

GGAGTGCCGT CCCCGGCCTT CTCGCGGCCG TGATGCACCT CCCTCTGCGG TGGGGTCCGG 60

GACATGGCAG GTAATGAGCC GGACGAGGGG AGCCAAGCTG GAGTTTACAC AGGCAAAGTG 120

55

EP 0 875 567 A2

	TCAGAAAAGA GTAGCCTGGG CTGTCTGGAA ATCTGAGCC ATG GAC TTT CCC CAG	174
	Met Asp Phe Pro Gln	
5	1 5	
	CAC AGC CAG CAT GTC TTG GAA CAG CTG AAC CAG CAG CGG CAG CTG GGG	222
	His Ser Gln His Val Leu Glu Gln Leu Asn Gln Gln Arg Gln Leu Gly	
	10 15 20	
10	CTT CTC TGT GAC TGC ACC TTT GTG GTG GAC GGT GTT CAC TTT AAG GCT	270
	Leu Leu Cys Asp Cys Thr Phe Val Val Asp Gly Val His Phe Lys Ala	
	25 30 35	
15	CAT AAA GCA GTG CTG GCG GCC TGC AGC GAG TAC TTC AAG ATG CTC TTC	318
	His Lys Ala Val Leu Ala Ala Cys Ser Glu Tyr Phe Lys Met Leu Phe	
	40 45 50	
20	GTG GAC CAG AAG GAC GTG GTG CAC CTG GAC ATC AGT AAC GCG GCA GGC	366
	Val Asp Gln Lys Asp Val Val His Leu Asp Ile Ser Asn Ala Ala Gly	
	55 60 65	
25	CTG GGG CAG ATG CTG GAG TTT ATG TAC ACG GCC AAG CTG AGC CTG AGC	414
	Leu Gly Gln Met Leu Glu Phe Met Tyr Thr Ala Lys Leu Ser Leu Ser	
	70 75 80 85	
30	CCT GAG AAC GTG GAT GAT GTG CTG GCC GTG GCC ACT TTC CTC CAA ATG	462
	Pro Glu Asn Val Asp Asp Val Leu Ala Val Ala Thr Phe Leu Gln Met	
	90 95 100	
35	CAG GAC ATC ATC ACG GCC TGC CAT GCC CTC AAG TCA CTT GCT GAG CCG	510
	Gln Asp Ile Ile Thr Ala Cys His Ala Leu Lys Ser Leu Ala Glu Pro	
	105 110 115	
40	GCT ACC AGC CCT GGG GGA AAT GCG GAG GCC TTG GCC ACA GAA GGA GGG	558
	Ala Thr Ser Pro Gly Gly Asn Ala Glu Ala Leu Ala Thr Glu Gly Gly	
	120 125 130	
45	GAC AAG AGA GCC AAA GAG GAG AAG GTG GCC ACC AGC ACG CTG AGC AGG	606
	Asp Lys Arg Ala Lys Glu Glu Lys Val Ala Thr Ser Thr Leu Ser Arg	
	135 140 145	
50	CTG GAG CAG GCA GGA CGC AGC ACA CCC ATA GGC CCC AGC AGG GAC CTC	654
	Leu Glu Gln Ala Gly Arg Ser Thr Pro Ile Gly Pro Ser Arg Asp Leu	
	150 155 160 165	
55	AAG GAG GAG CGC GGC GGT CAG GCC CAG AGT GCG GCC AGC GGT GCA GAG	702
	Lys Glu Glu Arg Gly Gly Gln Ala Gln Ser Ala Ala Ser Gly Ala Glu	
	170 175 180	
60	CAG ACA GAG AAA GCC GAT GCG CCC CGG GAG CCG CCG CCT GTG GAG CTC	750
	Gln Thr Glu Lys Ala Asp Ala Pro Arg Glu Pro Pro Pro Val Glu Leu	
	185 190 195	



EP 0 875 567 A2

	AAG CCA GAC CCC ACG AGT GGC ATG GCT GCC GCA GAA GCT GAG GCC GCT	798
	Lys Pro Asp Pro Thr Ser Gly Met Ala Ala Ala Glu Ala Glu Ala Ala	
	200 205 210	
5		
	TTG TCC GAG AGC TCG GAG CAA GAA ATG GAG GTG GAG CCC GCC CGG AAA	846
	Leu Ser Glu Ser Ser Glu Gln Glu Met Glu Val Glu Pro Ala Arg Lys	
	215 220 225	
10		
	GGG GAA GAG GAG CAA AAG GAG CAA GAG GAG CAA GAG GAG GAG GGC GCA	894
	Gly Glu Glu Glu Gln Lys Glu Gln Glu Glu Gln Glu Glu Glu Gly Ala	
	230 235 240 245	
15		
	GGG CCA GCT GAG GTC AAG GAG GAG GGT TCC CAG CTG GAG AAC GGA GAG	942
	Gly Pro Ala Glu Val Lys Glu Glu Gly Ser Gln Leu Glu Asn Gly Glu	
	250 255 260	
20		
	GCC CCC GAG GAG AAC GAG AAT GAG GAG TCA GCG GGC ACA GAC TCG GGG	990
	Ala Pro Glu Glu Asn Glu Asn Glu Glu Ser Ala Gly Thr Asp Ser Gly	
	265 270 275	
25		
	CAG GAG CTC GGC TCC GAG GCC CGG GGC CTG CGC TCA GGC ACC TAC GGC	1038
	Gln Glu Leu Gly Ser Glu Ala Arg Gly Leu Arg Ser Gly Thr Tyr Gly	
	280 285 290	
30		
	GAC CGC ACG GAG TCC AAG GCC TAC GGC TCC GTC ATC CAC AAG TGC GAG	1086
	Asp Arg Thr Glu Ser Lys Ala Tyr Gly Ser Val Ile His Lys Cys Glu	
	295 300 305	
35		
	GAC TGT GGG AAG GAG TTC ACG CAC ACG GGG AAC TTC AAG CGG CAC ATC	1134
	Asp Cys Gly Lys Glu Phe Thr His Thr Gly Asn Phe Lys Arg His Ile	
	310 315 320 325	
40		
	CGC ATC CAC ACG GGG GAG AAG CCC TTC TCG TGC CGG GAG TGC AGC AAG	1182
	Arg Ile His Thr Gly Glu Lys Pro Phe Ser Cys Arg Glu Cys Ser Lys	
	330 335 340	
45		
	GCC TTT TCC GAC CCG GCC GCG TGC AAG GCC CAT GAG AAG ACG CAC AGC	1230
	Ala Phe Ser Asp Pro Ala Ala Cys Lys Ala His Glu Lys Thr His Ser	
	345 350 355	
50		
	CCT CTG AAG CCC TAC GGC TGC GAG GAG TGC GGG AAG AGC TAC CGC CTC	1278
	Pro Leu Lys Pro Tyr Gly Cys Glu Glu Cys Gly Lys Ser Tyr Arg Leu	
	360 365 370	
55		
	ATC AGC CTG CTG AAC CTG CAC AAG AAG CGG CAC TCG GGC GAG GCG CGC	1326
	Ile Ser Leu Leu Asn Leu His Lys Lys Arg His Ser Gly Glu Ala Arg	
	375 380 385	
60		
	TAC CGC TGC GAG GAC TGC GGC AAG CTC TTC ACC ACC TCG GGC AAC CTC	1374
	Tyr Arg Cys Glu Asp Cys Gly Lys Leu Phe Thr Thr Ser Gly Asn Leu	
	390 395 400 405	

EP 0 875 567 A2

	AAG CGC CAC CAG CTG GTG CAC AGC GGC GAG AAG CCC TAC CAG TGC GAC Lys Arg His Gln Leu Val His Ser Gly Glu Lys Pro Tyr Gln Cys Asp 410 415 420	1422
5	TAC TGC GGC CGC TCC TTC TCC GAC CCC ACT TCC AAG ATG CGC CAC CTG Tyr Cys Gly Arg Ser Phe Ser Asp Pro Thr Ser Lys Met Arg His Leu 425 430 435	1470
10	GAG ACC CAC GAC ACG GAC AAG GAG CAC AAG TGC CCA CAC TGC GAC AAG Glu Thr His Asp Thr Asp Lys Glu His Lys Cys Pro His Cys Asp Lys 440 445 450	1518
15	AAG TTC AAC CAG GTA GGG AAC CTG AAG GCC CAC CTG AAG ATC CAC ATC Lys Phe Asn Gln Val Gly Asn Leu Lys Ala His Leu Lys Ile His Ile 455 460 465	1566
20	GCT GAC GGG CCC CTC AAG TGC CGA GAG TGT GGG AAG CAG TTC ACC ACC Ala Asp Gly Pro Leu Lys Cys Arg Glu Cys Gly Lys Gln Phe Thr Thr 470 475 480 485	1614
25	TCA GGG AAC CTG AAG CGG CAA CTT CGG ATC CAC AGC GGG GAG AAG CCC Ser Gly Asn Leu Lys Arg Gln Leu Arg Ile His Ser Gly Glu Lys Pro 490 495 500	1662
30	TAC GTG TGC ATC CAC TGC CAG CGA CAG TTT GCA GAC CCC GGC GCT CTG Tyr Val Cys Ile His Cys Gln Arg Gln Phe Ala Asp Pro Gly Ala Leu 505 510 515	1710
35	CAG CGG CAC GTC CGC ATT CAC ACA GGT GAG AAG CCA TGC CAG TGT GTG Gln Arg His Val Arg Ile His Thr Gly Glu Lys Pro Cys Gln Cys Val 520 525 530	1758
40	ATG TGC GGT AAG GCC TTC ACC CAG GCC AGC TCC CTC ATC GCC CAC GTG Met Cys Gly Lys Ala Phe Thr Gln Ala Ser Ser Leu Ile Ala His Val 535 540 545	1806
45	CGC CAG CAC ACC GGG GAG AAG CCC TAC GTC TGC GAG CGC TGC GGC AAG Arg Gln His Thr Gly Glu Lys Pro Tyr Val Cys Glu Arg Cys Gly Lys 550 555 560 565	1854
50	AGA TTC GTC CAG TCC AGC CAG TTG GCC AAT CAT ATT CGC CAC CAC GAC Arg Phe Val Gln Ser Ser Gln Leu Ala Asn His Ile Arg His His Asp 570 575 580	1902
55	AAC ATC CGC CCA CAC AAG TGC AGC GTG TGC AGC AAG GCC TTC GTG AAC Asn Ile Arg Pro His Lys Cys Ser Val Cys Ser Lys Ala Phe Val Asn 585 590 595	1950
60	GTG GGG GAC CTG TCC AAG CAC ATC ATC ATT CAC ACT GGA GAG AAG CCT Val Gly Asp Leu Ser Lys His Ile Ile Ile His Thr Gly Glu Lys Pro 600 605 610	1998



EP 0 875 567 A2

	TAC CTG TGT GAT AAG TGT GGG CGT GGC TTC AAC CGG GTA GAC AAC CTG	2046
	Tyr Leu Cys Asp Lys Cys Gly Arg Gly Phe Asn Arg Val Asp Asn Leu	
	615 620 625	
5	CGC TCC CAC GTG AAG ACC GTG CAC CAG GGC AAG GCA GGC ATC AAG ATC	2094
	Arg Ser His Val Lys Thr Val His Gln Gly Lys Ala Gly Ile Lys Ile	
	630 635 640 645	
10	CTG GAG CCC GAG GAG GGC AGT GAG GTC AGC GTG GTC ACT GTG GAT GAC	2142
	Leu Glu Pro Glu Glu Gly Ser Glu Val Ser Val Val Thr Val Asp Asp	
	650 655 660	
	ATG GTC ACG CTG GCT ACC GAG GCA CTG GCA GCG ACA GCC GTC ACT CAG	2190
15	Met Val Thr Leu Ala Thr Glu Ala Leu Ala Ala Thr Ala Val Thr Gln	
	665 670 675	
	CTC ACA GTG GTG CCG GTG GGA GCT GCA GTG ACA GCC GAT GAG ACG GAA	2238
	Leu Thr Val Val Pro Val Gly Ala Ala Val Thr Ala Asp Glu Thr Glu	
20	680 685 690	
	GTC CTG AAG GCC GAG ATC AGC AAA GCT GTG AAG CAA GTG CAG GAA GAA	2286
	Val Leu Lys Ala Glu Ile Ser Lys Ala Val Lys Gln Val Gln Glu Glu	
	695 700 705	
25	GAC CCC AAC ACT CAC ATC CTC TAC GCC TGT GAC TCC TGT GGG GAC AAG	2334
	Asp Pro Asn Thr His Ile Leu Tyr Ala Cys Asp Ser Cys Gly Asp Lys	
	710 715 720 725	
30	TTT CTG GAT GCC AAC AGC CTG GCT CAG CAT GTG CGA ATC CAC ACA GCC	2382
	Phe Leu Asp Ala Asn Ser Leu Ala Gln His Val Arg Ile His Thr Ala	
	730 735 740	
	CAG GCA CTG GTC ATG TTC CAG ACA GAC GCG GAC TTC TAT CAG CAG TAT	2430
35	Gln Ala Leu Val Met Phe Gln Thr Asp Ala Asp Phe Tyr Gln Gln Tyr	
	745 750 755	
	GGG CCA GGT GGC ACG TGG CCT GCC GGG CAG GTG CTG CAG GCT GGG GAG	2478
	Gly Pro Gly Gly Thr Trp Pro Ala Gly Gln Val Leu Gln Ala Gly Glu	
40	760 765 770	
	CTG GTC TTC CGC CCT CGC GAC GGG GCT GAG GGC CAG CCC GCA CTG GCA	2526
	Leu Val Phe Arg Pro Arg Asp Gly Ala Glu Gly Gln Pro Ala Leu Ala	
	775 780 785	
45	GAG ACC TCC CCT ACA CCT CCT GAA TGT CCC CCG CCT GCC GAG TGAGCTGGCG	2578
	Glu Thr Ser Pro Thr Pro Pro Glu Cys Pro Pro Pro Ala Glu	
	790 795 800	
	GCCCTTCTGA CTGTTTATTT AAGGATGGAT GGCACCCTGG AACCGGGAAG GGTGGCCTGT	2638
50	TCCCTAGAGA GAATAAATTG GATTATTTTC TAAAAAAA AA	2680

(2) INFORMATION ZU SEQ ID NO: 2:

## (i) SEQUENZ CHARAKTERISTIKA:

(A) LÄNGE: 803 Aminosäuren

(B) ART: Aminosäure

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 2:

Met Asp Phe Pro Gln His Ser Gln His Val Leu Glu Gln Leu Asn Gln  
 1 5 10 15  
 Gln Arg Gln Leu Gly Leu Leu Cys Asp Cys Thr Phe Val Val Asp Gly  
 20 25 30  
 Val His Phe Lys Ala His Lys Ala Val Leu Ala Ala Cys Ser Glu Tyr  
 35 40 45  
 Phe Lys Met Leu Phe Val Asp Gln Lys Asp Val Val His Leu Asp Ile  
 50 55 60  
 Ser Asn Ala Ala Gly Leu Gly Gln Met Leu Glu Phe Met Tyr Thr Ala  
 65 70 75 80  
 Lys Leu Ser Leu Ser Pro Glu Asn Val Asp Asp Val Leu Ala Val Ala  
 85 90 95  
 Thr Phe Leu Gln Met Gln Asp Ile Ile Thr Ala Cys His Ala Leu Lys  
 100 105 110  
 Ser Leu Ala Glu Pro Ala Thr Ser Pro Gly Gly Asn Ala Glu Ala Leu  
 115 120 125  
 Ala Thr Glu Gly Gly Asp Lys Arg Ala Lys Glu Glu Lys Val Ala Thr  
 130 135 140  
 Ser Thr Leu Ser Arg Leu Glu Gln Ala Gly Arg Ser Thr Pro Ile Gly  
 145 150 155 160  
 Pro Ser Arg Asp Leu Lys Glu Glu Arg Gly Gly Gln Ala Gln Ser Ala  
 165 170 175  
 Ala Ser Gly Ala Glu Gln Thr Glu Lys Ala Asp Ala Pro Arg Glu Pro  
 180 185 190  
 Pro Pro Val Glu Leu Lys Pro Asp Pro Thr Ser Gly Met Ala Ala Ala  
 195 200 205  
 Glu Ala Glu Ala Ala Leu Ser Glu Ser Ser Glu Gln Glu Met Glu Val  
 210 215 220  
 Glu Pro Ala Arg Lys Gly Glu Glu Glu Gln Lys Glu Gln Glu Glu Gln  
 225 230 235 240

EP 0 875 567 A2

Pro Cys Gln Cys Val Met Cys Gly Lys Ala Phe Thr Gln Ala Ser Ser  
530 535 540

5 Leu Ile Ala His Val Arg Gln His Thr Gly Glu Lys Pro Tyr Val Cys  
545 550 555 560

Glu Arg Cys Gly Lys Arg Phe Val Gln Ser Ser Gln Leu Ala Asn His  
565 570 575

10 Ile Arg His His Asp Asn Ile Arg Pro His Lys Cys Ser Val Cys Ser  
580 585 590

Lys Ala Phe Val Asn Val Gly Asp Leu Ser Lys His Ile Ile Ile His  
595 600 605

15 Thr Gly Glu Lys Pro Tyr Leu Cys Asp Lys Cys Gly Arg Gly Phe Asn  
610 615 620

20 Arg Val Asp Asn Leu Arg Ser His Val Lys Thr Val His Gln Gly Lys  
625 630 635 640

Ala Gly Ile Lys Ile Leu Glu Pro Glu Glu Gly Ser Glu Val Ser Val  
645 650 655

25 Val Thr Val Asp Asp Met Val Thr Leu Ala Thr Glu Ala Leu Ala Ala  
660 665 670

Thr Ala Val Thr Gln Leu Thr Val Val Pro Val Gly Ala Ala Val Thr  
675 680 685

30 Ala Asp Glu Thr Glu Val Leu Lys Ala Glu Ile Ser Lys Ala Val Lys  
690 695 700

Gln Val Gln Glu Glu Asp Pro Asn Thr His Ile Leu Tyr Ala Cys Asp  
705 710 715 720

Ser Cys Gly Asp Lys Phe Leu Asp Ala Asn Ser Leu Ala Gln His Val  
725 730 735

40 Arg Ile His Thr Ala Gln Ala Leu Val Met Phe Gln Thr Asp Ala Asp  
740 745 750

Phe Tyr Gln Gln Tyr Gly Pro Gly Gly Thr Trp Pro Ala Gly Gln Val  
755 760 765

45 Leu Gln Ala Gly Glu Leu Val Phe Arg Pro Arg Asp Gly Ala Glu Gly  
770 775 780

Gln Pro Ala Leu Ala Glu Thr Ser Pro Thr Pro Pro Glu Cys Pro Pro  
785 790 795 800

Pro Ala Glu

55

EP 0 875 567 A2

	Glu Glu Glu Gly Ala Gly Pro Ala Glu Val Lys Glu Glu Gly Ser Gln	
	245	250 255
5	Leu Glu Asn Gly Glu Ala Pro Glu Glu Asn Glu Asn Glu Glu Ser Ala	
	260	265 270
	Gly Thr Asp Ser Gly Gln Glu Leu Gly Ser Glu Ala Arg Gly Leu Arg	
	275	280 285
10	Ser Gly Thr Tyr Gly Asp Arg Thr Glu Ser Lys Ala Tyr Gly Ser Val	
	290	295 300
	Ile His Lys Cys Glu Asp Cys Gly Lys Glu Phe Thr His Thr Gly Asn	
15	305	310 315 320
	Phe Lys Arg His Ile Arg Ile His Thr Gly Glu Lys Pro Phe Ser Cys	
	325	330 335
20	Arg Glu Cys Ser Lys Ala Phe Ser Asp Pro Ala Ala Cys Lys Ala His	
	340	345 350
	Glu Lys Thr His Ser Pro Leu Lys Pro Tyr Gly Cys Glu Glu Cys Gly	
	355	360 365
25	Lys Ser Tyr Arg Leu Ile Ser Leu Leu Asn Leu His Lys Lys Arg His	
	370	375 380
	Ser Gly Glu Ala Arg Tyr Arg Cys Glu Asp Cys Gly Lys Leu Phe Thr	
30	385	390 395 400
	Thr Ser Gly Asn Leu Lys Arg His Gln Leu Val His Ser Gly Glu Lys	
	405	410 415
35	Pro Tyr Gln Cys Asp Tyr Cys Gly Arg Ser Phe Ser Asp Pro Thr Ser	
	420	425 430
	Lys Met Arg His Leu Glu Thr His Asp Thr Asp Lys Glu His Lys Cys	
	435	440 445
40	Pro His Cys Asp Lys Lys Phe Asn Gln Val Gly Asn Leu Lys Ala His	
	450	455 460
	Leu Lys Ile His Ile Ala Asp Gly Pro Leu Lys Cys Arg Glu Cys Gly	
45	465	470 475 480
	Lys Gln Phe Thr Thr Ser Gly Asn Leu Lys Arg Gln Leu Arg Ile His	
	485	490 495
50	Ser Gly Glu Lys Pro Tyr Val Cys Ile His Cys Gln Arg Gln Phe Ala	
	500	505 510
	Asp Pro Gly Ala Leu Gln Arg His Val Arg Ile His Thr Gly Glu Lys	
	515	520 525

55

Patentansprüche

- 5 1. Isoliertes Protein mit der in SEQ ID NO:2 dargestellten Aminosäuresequenz sowie die daraus durch Substitution, Insertion oder Deletion von einem oder mehreren Aminosäureresten erhältlichen Muteine, die noch die wesentlichen biologischen Eigenschaften des in SEQ ID NO:2 dargestellten Proteins besitzen.
2. Protein gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um ein humanes Protein handelt.
- 10 3. Nukleinsäuresequenz codierend für ein Protein gemäß Anspruch 1.
4. Nukleinsäuresequenz nach Anspruch 3, dadurch gekennzeichnet, daß sie für ein Protein codiert, das mindestens 95 % Identität mit der in SEQ ID NO:2 dargestellten Sequenz besitzt.
- 15 5. Nukleinsäuresequenz nach Anspruch 3, dadurch gekennzeichnet, daß sie die in SEQ ID NO:1 dargestellte Struktur besitzt.
6. Vektor enthaltend eine Nukleinsäuresequenz gemäß Anspruch 3 - 5, funktionell verknüpft mit mindestens einem Regulationselement.
- 20 7. Wirtsorganismus, transformiert mit einer Nukleinsäuresequenz nach Anspruch 3.
8. Wirtsorganismus, transformiert mit einem Vektor gemäß Anspruch 6.
- 25 9. Verfahren zur Herstellung eines Proteins gemäß Anspruch 1, dadurch gekennzeichnet, daß man einen Wirtsorganismus gemäß Anspruch 6 unter Bedingungen kultiviert, die die Expression des Proteins erlauben und anschließend das exprimierte Protein vom Wirtsorganismus abtrennt und in reiner Form isoliert.
- 30 10. Verwendung eines Proteins gemäß Anspruch 1 zur Identifizierung von spezifischen transkriptionsmodulierenden Substanzen.
11. Verfahren zur Identifizierung von spezifischen transkriptionsmodulierenden Substanzen, das folgende Schritte umfaßt:
  - 35 (a) Inkubation des Proteins gemäß Anspruch 1 mit dem Genprodukt von myc unter Bedingungen, unter denen sich ein Proteinkomplex zwischen diesen beiden Proteinen ausbildet,
  - (b) Inkubation der beiden Proteine unter ansonst gleichen Bedingungen wie (a) jedoch in Anwesenheit einer oder mehrerer Substanzen, die auf spezifische transkriptionsmodulierende Aktivitäten zu testen sind,
  - 40 (c) Ermitteln des Unterschiedes in der Proteinkomplexbildung zwischen (b) und (a),
  - (d) Auswahl solcher Substanzen, bei denen gemäß Schritt (b) eine andere Proteinkomplexbildung erhalten wurde als bei Schritt (a).
- 45 12. Verwendung eines Proteins gemäß Anspruch 1 als Antigen zur Herstellung von spezifischen Antikörpern.
13. Verwendung einer Nukleinsäuresequenz nach Anspruch 3 zur Gentherapie.
14. Verwendung einer zu der Sequenz gemäß Anspruch 3 komplementären Nukleinsäuresequenz zur Gentherapie.
- 50 15. Verwendung nach Anspruch 13 oder 14, dadurch gekennzeichnet, daß man durch die exogen zugeführte Nukleinsäuresequenz die zelluläre Konzentration des Proteins gemäß Anspruch 1 erhöht oder erniedrigt.

55